連立方程式の数値解の一つの試み (数値解析の基礎理論)

松井, 俊二; 平野, 菅保

数理解析研究所講究録
107巻 1971年 63-77

連立方程式の数値解の一つの試み

東芝　松井　俊二
東芝　平野　菅保

多元連立1次方程式を解く場合、解が常識にあわないと、計算手順を変えたり、又は計算桁数を増加させて計算し、それでもなお解が常識にあわないと、係数行列が悪条件なのであるとする傾向がある。しかし、眞に係数行列が悪条件であるのか、悪条件とはどんなことであるのか、等をはっきりさせる必要がある。

例えば、係数行列の絶対値最大の固有値 λ_{max} と絶対値最小の固有値 λ_{min} の比が $|\lambda_{max}/\lambda_{min}| \gg 1$ であれば、悪条件であるとすることが一般に行われているが、同じ係数行列でも定数項ベクトルの数値によっては解の精度がよいときもあるし、行又は列に適当な定数を乗ずると、絶対値最大の固有値 λ_{max} と絶対値最小の固有値 λ_{min} は変化し、したがって比 $|\lambda_{max}/\lambda_{min}|$ が小さくなることがある。これで係数行列の悪条件が全て解決できるであろうか。それはできない

。係数行列の悪条件とはこのできない場合を指すべきではないか。しかし、固有値、固有ベクトル等を用いて悪条件である係数行列を示せば、思考上便利である。一方実際の計算では、係数行列の固有値、固有ベクトルを知らずに計算をすることが一般である。したがつて固有値、固有ベクトルを知らなければ、係数行列が悪条件か、否かがわからないようでは困るのであつて、求解計算をしていながら、係数行列が悪条件か、否かがわからないものであろうか。

数値計算は有限桁の実数で行われるので、誤差が必ず導入されるし、又数値計算で誤差が導入されないとしても、計算に際して与えられる数値は一般に誤差を含んでいるから、その誤差の影響によつて解の1桁目まで誤差が入つている場合もある。すなわち、係数行列、定数ベクトルの各要素が全てn桁の精度をもつていても、解の精度は、どんな計算方法を用いても、1桁もないことがある。このような場合には、計算桁数を増加して求解計算をしても何ら意味がない。しかし、解法の計算手順が好ましくなく、求解計算の途中で誤差を導入してしまうときには、数値計算による誤差をなるべく入れないようにするため、計算桁数を増加することは意味がある。ここでは、一般の多元連立方程式を反復法で解く場合、連立1次方程式の特徴に、昨年述べた代数方程式の数値解法

を加味して考えると、何らかの役に立つであろうことを期待して、ガウスの消去法を中心にして多元連立1次方程式について述べてみたい。

I 数式の誤差と解の誤差

多元連立1次方程式を解くとき、計算桁数は有限桁で行うので、係数として採用した桁より下位の桁は誤差になっている。したがって数値計算による誤差が入らなくても、与えられる係数の誤差により、解は誤差を含んでしまう。

いま、与えられた多元連立1次方程式を次式とする。

$$\sum_{j=1}^{n}(a_{ij}\pm\Delta a_{ij})x_j-(K_i\pm\Delta K_i)=0 \qquad (i=1,2,\cdots,n)$$

$\pm\Delta a_{ij}$: a_{ij} の誤差，　$\pm\Delta K_i$: K_i の誤差　(1)

真の解 x_{Tj} を(1)式に代入すると、

$$\sum_{j=1}^{n}[(a_{ij}\pm\Delta a_{ij})x_{Tj}\mp\Delta a_{ij}x_{Tj}]-(K_i\pm\Delta K_i)\mp\Delta K_i=0$$

$$(i=1,2,\cdots,n) \qquad (2)$$

各項に誤差項を入れなくては等号が成立しない。次に(2)式の誤差項の中で絶対値最大の数値 ε_i を求める。

$$\varepsilon_i=\max_{j=1,2,\cdots,n}(|\mp\Delta a_{ij}x_{Tj}|,|\mp\Delta K_i|)$$

$$(i=1,2,\cdots,n) \qquad (3)$$

(3)式で得られる ε_i を i 式の数式誤差とする。ε_i はどのような数値を意味するか。(2)式を変形してみる。

$$(a_{ik}\pm\Delta a_{ik})x_{Tk}\mp\Delta a_{ik}x_{Tk}=(K_i\pm\Delta K_i)\mp\Delta K_i-\sum_{j=1}^{n'}[(a_{ij}\pm\Delta a_{ij})x_{Tj}\mp\Delta a_{ij}x_{Tj}]$$

n'：$j=k$ を除く　　$(i=1,2,\cdots\cdots,n)$　　(4)

(4)式の右辺に含まれる解 x_{Tj} ($j=k$ を除く)が、たとえ誤差を含んでいないとしても．(4)式には誤差 ε_i を左辺又は右辺に含んでいるから．(4)式から解 x_{Tk} を求めたとすると．解 x_{Tk} の誤差 $\pm\Delta x_{Tk}$ は

$$\pm\Delta x_{Tk}=|\varepsilon_i/(a_{ik}\pm\Delta a_{ik})| \qquad (5)$$

で求められる。したがつて解 x_{Tk} の相対誤差は

$$|\pm\Delta x_{Tk}/x_{Tk}|=|\varepsilon_i/(a_{ik}\pm\Delta a_{ik})x_{Tk}| \qquad (6)$$

で求められる。それでは、i 式からはどの解を求めたならばよいのか。既知の解を(4)式の右辺に代入して合計する場合．桁落ち現象をおこして有効桁数を減少させないようにするためには

$$C_i=|(a_{ik}\pm\Delta a_{ik})x_{Tk}|=\max_{j=1,2,\cdots,n}(|(a_{ij}\pm\Delta a_{ij})x_{Tj}|)$$

$(i=1,2,\cdots\cdots,n)$　　(7)

(7)式で選ばれた解 x_{Tk} を求めればよい。これは又(6)式の相対誤差を小さくするような解 x_{Tk} を選んでいることにもなる。(7)式の C_i は i 式の絶対値を表わすとする。

次に、各数式の誤差を等しくするために(2)式を ε_i で除す。

$$\sum_{j=1}^{n}(a_{ij}\pm\Delta a_{ij})x_{Tj}/\varepsilon_i-(K_i\pm\Delta K_i)/\varepsilon_i=0$$

$$\varepsilon_i^{(0)}=\varepsilon_i/\varepsilon_i\ ,\quad C_i^{(0)}=C_i/\varepsilon_i\quad(i=1,2,\cdots\cdots,n)\qquad(8)$$

(8)式では数式の誤差がどの式も1.0になっている。又1.0より小さい誤差項は式より除いた。(8)式の係数と定数を書きかえて

$$\begin{aligned}b_{ij}^{(0)}&=(a_{ij}\pm\Delta a_{ij})/\varepsilon_i\\ L_i^{(0)}&=(K_i\pm\Delta K_i)/\varepsilon_i\end{aligned}\quad\begin{pmatrix}i=1,2,\cdots\cdots,n\\ j=1,2,\cdots\cdots,n\end{pmatrix}\qquad(9)$$

$$\sum_{j=1}^{n}b_{ij}^{(0)}x_{Tj}-L_i^{(0)}=0\qquad(i=1,2,\cdots\cdots,n)\qquad(10)$$

(10)式で表わす。(10)式よりガウスの消去法を行う。m回目の消去段階で、ピボットとしては

$$\begin{aligned}|b_{kl}^{(m-1)}x_{Tl}|&=\max_{i=\prime\ j=\prime}(|b_{ij}^{(m-1)}x_{Tj}|)=\max_{i=\prime}(|C_i^{(m-1)}/\varepsilon_i^{(m-1)}|)\\ &=\max_{i=\prime}(|C_i^{(m-1)}|)=|C_k^{(m-1)}|\end{aligned}\qquad(11)$$

$i=\prime$ は$(m-1)$回までに消去に用いられなかつた数式番号

$j=\prime$ は$(m-1)$回までに消去に用いられなかつた変数番号

(11)式で求めたkl要素を用いて、k式で消去する。

$$\begin{aligned}&\left(\sum_j{}'b_{i'j}^{(m-1)}x_{Tj}-L_{i'}^{(m-1)}\right)-(b_{i'l}^{(m-1)}/b_{kl}^{(m-1)})\left(\sum_j{}'b_{kj}^{(m-1)}x_{Tj}-L_k^{(m-1)}\right)\\ &=\sum_j{}''b_{i'j}^{(m)}x_{Tj}-L_{i'}^{(m)}\qquad(i'=\prime\prime)\end{aligned}\qquad(12)$$

$$(b_{i'l}^{(m-1)}x_{Tl}/b_{kl}^{(m-1)}x_{Tl})=(b_{i'l}^{(m-1)}/b_{kl}^{(m-1)})\leqq 1.0\qquad(13)$$

$\sum_j{}'$ は $j=\prime$ と同じ変数について

$\sum_j{}''$ は $j=\prime$ からl変数を除いた変数について

$i'=\prime\prime$ は $i=\prime$ からk式を除いた数式について

i'式の絶対値 $C_{i'}^{(m)}$ は

$$C_{i'}^{(m)} = \max_{j=''} (| b_{i'j}^{(m)} x_{Tj} |) \qquad (i' = '') \qquad (14)$$

$j=''$ は $j=1$ から l 変数を除いた変数について

i'式の誤差は2通り考えられる．まず

$$\varepsilon_{i'}^{(m)} = | \varepsilon_{i'}^{(m-1)} | + | (b_{i'l}^{(m-1)} / b_{kl}^{(m-1)}) \varepsilon_{k}^{(m-1)} | \qquad (15)$$

がある．これは（12）式のいずれかの項が（15）式で表わされる誤差 $\varepsilon_{i'}^{(m)}$ を含んでいることになる．しかし，これは i'式の誤差 $\varepsilon_{i'}^{(m-1)}$ と k式の誤差 $\varepsilon_{k}^{(m-1)}$ が共に同じ変数の項による誤差であって，しかも（12）式で増加するようになったときである．したがって次式を用いたほうが実際にはよく一致する．

$$\varepsilon_{i'}^{(m)} = \max (| \varepsilon_{i'}^{(m-1)} |, | (b_{i'l}^{(m-1)} / b_{kl}^{(m-1)}) \varepsilon_{k}^{(m-1)} |) = 1.0 \qquad (16)$$

ここで注意すべきことは，（15），（16）式共に $b_{kl}^{(m-1)}$ による除算，$b_{i'l}^{(m-1)}$ による乗算では誤差は入らないことである．

順次消去して，最後に変数が1つになったとき

$$b_{k'l'}^{(n-1)} x_{Tl'} = L_{k'}^{(n-1)} \qquad \varepsilon_{k'}^{(n-1)} = 1.0 \qquad (17)$$

であるから，$x_{Tl'}$ の相対誤差は

$$| \pm \Delta x_{Tl'} / x_{Tl'} | = | \varepsilon_{k'}^{(n-1)} / b_{k'l'}^{(n-1)} x_{Tl'} | = | 1.0 / b_{k'l'}^{(n-1)} x_{Tl'} | \qquad (18)$$

で求められる．ところが，実際には，眞の解 x_{Tj} は未知であるから近似解 $x_{Tj} \pm \Delta x_{Tj}$ を x_{Tj} の代りに用いる．したがって

近似解の精度が悪いと、(17)式の左辺と右辺が1.0^*より上位の桁で異なつてくる。その場合は、更に近似解の精度をよくするために

$$\bar{x}_{Tl'}=[L_{k'}^{(m-1)}/\{b_{k'l'}^{(m-1)}(x_{Tl'}\pm\Delta x_{Tl'})\}](x_{Tl'}\pm\Delta x_{Tl'}) \qquad (19)$$

(19)式により得られる$\bar{x}_{Tl'}$を近似解として採用する。

(19)式により近似解を補正し、他の解も順次近似値を次のように補正する。

$$b_{k''l''}^{(m)}x_{Tl''}=L_{k''}^{(m)}-\sum_{j=l'''}b_{k''}^{(m)}x_{Tj} \qquad (k''=1,2,\cdots\cdots,n) \qquad (20)$$

$\sum_{j=l'''}$ は解が既知の変数について

(20)式の左辺の数値と右辺の数値が1.0^*の桁より上位の桁で異なつているときは、(19)式と同様に

$$\bar{x}_{Tl''}=[\{L_{k''}^{(m)}-\sum_{j=l'''}b_{k''j}^{(m)}x_{Tj}\}/\{b_{k''l''}^{(m)}(x_{Tl''}\pm\Delta x_{Tl''})\}]$$
$$\times(x_{Tl''}\pm\Delta x_{Tl''}) \qquad (21)$$

(21)式で近似解の補正をする。次に、(1)，(20)式を消去に用いた順番に、式，変数共に並べかえ、得られた近似解を代入する。このとき、(19)式で近似解の補正を行つていれば、(20)式の並べかえた各式の左辺の値と右辺の値の差は計算桁数以下の桁である。(勿論1.0^*より下位の桁である。)いま、その差をそれぞれε_i($i=1,2,\cdots\cdots,n$)とすると、(1)式の並べかえた式ではどのようになるだろうか。各式の左辺の合計はゼロにならずに

$$\bar{\bar{\varepsilon}}_i = \sum_{k=1}^{i} \left(b_{ik}^{(k-1)} / b_{kk}^{(k-1)} \right) \bar{\varepsilon}_k \qquad (i = 1, 2, \cdots\cdots, n) \tag{22}$$

となるから、（13）式を考慮すると、一般に

$$\bar{\bar{\varepsilon}}_i \fallingdotseq 1.0 \tag{23}$$

となる。

計算桁数は（10）式で考えると、小数点以下数桁まであればよく、それ以上は無駄な桁である。又ピボットとして選ばれる要素の順番と計算桁数が同じであれば、計算結果は等しい数値が得られるから、（10）式以後ピボットとして選ばれた要素の順番に、近似解が未知の場合でもピボットを選んでいれば、（10）式で決定できる計算桁数と同じ計算桁数で、可能な解の精度は得られる。

（17）式で得られる解 $x_{T\ell'}$ 以外の解の誤差はどうであるか。（20）式の右辺にある変数の解 x_{Tj} は相対誤差も既知であるとすると、（20）式の右辺各項の誤差の絶対値がわかる。この絶対値が全て 1.0 より小さければ、（20）式の数式誤差そのものが 1.0 であるから、$x_{T\ell''}$ の相対誤差は

$$\left| \pm \Delta x_{T\ell''} / x_{T\ell''} \right| = \left\{ L_{k''}^{(m)} - \sum_{j} b_{k''\ell''}^{(m)} x_{Tj} \right\} / \left\{ b_{k''\ell''}^{(m)} \left(x_{T\ell''} \pm \Delta x_{T\ell''} \right) \right\} \tag{24}$$

で求められるが、絶対値が 1.0 より大きい誤差項があるときは、誤差の打ち消しが行われることがあるので、（24）式は

使用できない。この場合は（10）式にもどつて、計算をしなおす。すなわち、各消去毎に消去に用いてよい式について、相対誤差が既知である解とその係数を乗じて各項の絶対誤差を計算し、その値がその項の含まれる i 式の数式誤差 1.0 より大きい項がある場合は、その大きい項の誤差 δ_i をその式の数式誤差として、δ_i の絶対値で i 式の係数、定数を除して i 式の数式誤差を 1.0 とする。しかし、その回の消去でピボット** として係数を選ばれなかつた式については、消去の終つた段階で δ_i を係数、定数に乗じておく。（これは、誤差の打消しが行われることがあるからである。）このようにして、相対誤差未知の解が 1 つになつた式が得られたならば（そのような式は 1 つとは限らないが、相対誤差の最も小さくなる式を用いる。）（18）式によりその解の相対誤差を求める。

注 * 1.0 の桁より上位の桁とは限らず、（22）式の絶対値を小さくするために、小数点以下数桁を一致させるほうがよい。

** 相対誤差が既知の解の係数はピボットとして選ばぬ。

Ⅱ 近似解が未知の場合と係数行列の悪条件

係数行列の絶対値最大の固有値 λ_{max} と絶対値最小の固有

値 λ_{min} の比の絶対値が

$$|\lambda_{max}/\lambda_{min}| \gg 1.0 \qquad (25)$$

であれば、係数行列が悪条件であるといわれているが、係数行列は一定の行、又は一定の列に絶対値が 1.0 より非常に小さい数値を乗ずると、(25)式の左辺はいくらでも大きくなる。逆に、係数行列の各要素の絶対値が(ゼロを含まぬ。)桁違いに異ならないように、各行，各列に定数を乗じて係数行列を作ると、(25)式の左辺の値が小さくなる場合がある。与えられた多元連立1次方程式を次式とする。

$$\sum_{j=1}^{n} a_{ij} x_j = K_i \qquad (i=1,2,\cdots\cdots,n) \qquad (26)$$

$$A_i = \max_{j=1,2,\cdots\cdots,n} (|a_{ij}|) \qquad (i=1,2,\cdots\cdots,n) \qquad (27)$$

$$\bar{A}_j = \max_{i=1,2,\cdots\cdots,n} (|a_{ij}/A_i|) \qquad (i=1,2,\cdots\cdots,n) \qquad (28)$$

(27)，(28)式の A_i，$\bar{A}_j$ を用いて(26)式を書きかえて

$$\left.\begin{aligned} a_{ij}^{(0)} &= a_{ij}/(A_i \cdot \bar{A}_j) \\ \bar{x}_j &= \bar{A}_j \cdot x_j \\ K_i^{(0)} &= K_i/A_i \end{aligned}\right. \qquad \begin{pmatrix} i=1,2,\cdots\cdots,n \\ j=1,2,\cdots\cdots,n \end{pmatrix} \qquad (29)$$

$$\sum_{j=1}^{n} a_{ij}^{(0)} \bar{x}_j = K_i^{(0)} \qquad (i=1,2,\cdots\cdots,n) \qquad (30)$$

(30)式からの計算法は「I　数式の誤差と解の誤差」で述べたように、消去に用いてよい式の係数中で絶対値最大の数値をピボットに選んで消去する。係数行列が悪条件の場合はピボットの数値の絶対値が段々小さくなり、(30)式で考え

れば誤差である．絶対値の小さな数値になる．したがって計算桁数を一定にして計算していると，（22）式の $\bar{\varepsilon}_i$ は小さくなり，この方法で得られた解を（30）式に代入しても，$\bar{\bar{\varepsilon}}_i$ の絶対値（左辺の合計と右辺との差）は大きくならず，係数行列が悪条件でないときと同様になる。又各式の絶対値最大の項，数式の絶対値 C_i と $\bar{\bar{\varepsilon}}_i$ との比 $|C_i/\bar{\bar{\varepsilon}}_i|$ が大きければ，式をよく満足しているが，C_i の絶対値は解の存在の仕方，すなわち定数項の数値によって異なるので，各式の誤差 $\bar{\bar{\varepsilon}}_i$ は小さくすることができても，C_i は解が未知の場合は大きくする方法がない。

（30）式より消去するときと，（10）式より消去するときとでは選ばれるピボットの順番が一般に異なるが，同じ場合には数式の絶対値 C_i は好ましい値をとってくれる。又選ばれるピボットの順番が好ましくないと，実質的には計算桁数の減少を意味し，逆に計算桁数を多くすると，ピボット選択を最適に行わなくても誤差は打ち消しで消える。

係数行列が悪条件である場合，各要素に適当に誤差を入れて逆行列を求めてみる。まず，絶対値の大きな誤差を入れて逆行列を求めると，その要素は1桁目から異なるが，入れる誤差を段々小さくして逆行列を求めると，順次1桁目，2桁目とその要素は正しい値になる．すなわち，誤差が小さくな

れば、悪条件でない係数行列の場合と同様に、誤差が1桁小さくなれば逆行列の要素も1桁精度がよくなる。

(30)式より、消去に用いてよい式の係数中絶対値最大の数値をピボットにして消去しているうち、ピボットの数値が(30)式より考えて誤差の数値になったならば、ピボットの行，列より先の、すなわち既に消去に用いた行，列を用いて他の列，行を多元連立1次方程式の定数項として比例定数を求め、その比例定数を(30)式に代入し、「I 数式の誤差と解の誤差」と同様に、数式誤差の範囲内で従属であるかを調べる必要がある。

一般にいわれる多元連立1次方程式における不定問題，不能問題は、数値計算ではどのようになるか。(1)式で考えるに、不定問題では係数行列のランクがm小さくなっているならば、m個の解が任意に決定でき、それに応じて他の解を決定すれば、(1)式がよく満足されている。しかし数値計算では、係数の誤差の影響によりランクが完全には落ちない。したがってそのランクの落ち方に応じてm個の解の下位の桁を任意に決定できる。不能問題では(1)式の定数項がゼロ、すなわち(1)式の左辺に得られた解を代入したとき、左辺の絶対値最大の項の誤差の桁程度に、定数項の絶対値がなるような絶対値の大きな解が得られる。

最後に、係数行列が悪条件の場合でも、例えば行列式の値が1～2桁の精度しかない場合でも、解の中には3桁以上の精度をもつものがあるときがあることをつけ加える。

Ⅲ　反復法

多元連立1次方程式を解くとき、ゼロの要素が多く、かつ計算桁数が少なければ、計算は楽である。そこで、ガウス・ザイデル法等の反復法が考えられる。係数行列が悪条件の場合、ガウスの消去法を用いて得られた精度の悪い解を、与えられた多元連立1次方程式に計算桁数を増加して代入し、左辺と右辺との差を求め、その差を定数項として2度目の解、すなわち解の補正項を求め、1回目の解に加えると、その解の精度は1回目に求めた解の精度よりよくなる。ただしこの場合、係数行列の要素の精度は計算桁数を増加して計算したときの桁までなければならない。すなわち、消去に用いなかつた係数行列の要素の下位の桁が左辺と右辺の差を求めるときの計算桁数の増加によつて解に影響してくる。

消去するとき、ピボットの選択が適当でないと、ガウスの消去法での計算桁数を減少させたことと同じ結果になるから、1度目に得られた解を(1)式に代入し、左辺と右辺の差を求める計算で計算桁数を増加させなくても、ガウスの消去法

を行ったときの実質的計算桁数よりも計算桁数を増加したことになる。したがって1回目と同じ順番でピボット選択してガウスの消去法により解の補正項を求めても、解の精度はよくなる。なおこの場合、補正項を求める2,3,…回目のガウスの消去法の計算では、誤差間の打ち消しが行われる。

次に、反復法を使用できるときの条件を考える。

$AX = B$ 　　A：係数行列

X：変数ベクトル

B：定数ベクトル　(31)

係数行列Aの代りに簡略係数行列$\tilde{A}$を用いて計算する。

$$\tilde{A}X = B \qquad (32)$$

(32)式より得られる解を(31)式に代入し、右辺より左辺を減じて差を求める。その差を定数ベクトルとして(32)式に代入し、解の補正項を求め、解を補正する。補正された解を(31)式に代入し、右辺より左辺を減じて差を求める。この反復により右辺と左辺の差がゼロに収束すれば、(31)式の解が求められる。i回目の補正項を求めるときの(32)式は$\tilde{A}X_i = B_i$　(33)

となり、(31)式に代入して右辺より左辺を減ずると

$$B_{i+1} = B_i - AX_i = \tilde{A}X_i - AX_i = (\tilde{A} - A)X_i \qquad (34)$$

$\tilde{A}$の固有値をλ_j，固有ベクトルをV_j***とすると

$$B_i = \sum_{j=1}^{n} C_{ij} V_j \qquad C_{ij}：B_i\text{の固有ベクトル}V_j\text{の成分} \tag{35}$$

$$X_i = \sum_{j=1}^{n} (C_{ij}/\lambda_j) V_j \tag{36}$$

となる。又 $\tilde{A}-A$ の固有値 μ_j，固有ベクトル W_j*** により

$$X_i = \sum_{j=1}^{n} d_{ij} W_j \qquad d_{ij}：X_i\text{の固有ベクトル}W_j\text{の成分} \tag{37}$$

$$B_{i+1} = \sum_{j=1}^{n} (d_{ij} \mu_j W_j) \tag{38}$$

と表わされる。固有値 λ_j（$j=1, 2, \cdots, n$）の中で絶対値最小の固有値を λ_{min} とすれば、（35），（36）式より

$$|X_i|^2 = \sum_{j=1}^{n} (C_{ij}/\lambda_j)^2 \leqq \left\{ \sum_{j=1}^{n} (C_{ij})^2 \right\} / \lambda_{min}^2 = |B_i|^2/\lambda_{min}^2 \tag{39}$$

（37），（38）式より、固有値 μ_j（$j=1, 2, \cdots, n$）の中で絶対値最大の固有値を μ_{max} とすれば

$$|B_{i+1}|^2 = \sum_{j=1}^{n} (d_{ij} \mu_j)^2 \leqq \left\{ \sum_{j=1}^{n} (d_{ij})^2 \right\} \mu_{max}^2 = |X_i|^2 \mu_{max}^2 \tag{40}$$

（39），（40）より

$$|B_{i+1}|^2 \leqq \mu_{max}^2 |X_i|^2 \leqq (\mu_{max}/\lambda_{min})^2 |B_i|^2 \tag{41}$$

したがって

$$(\mu_{max}/\lambda_{min})^2 < 1.0 \tag{42}$$

であれば、解は収束する。$\lambda_{min} = 0$ であると、実際の計算では、i の少ない間は i の増加にしたがって不定問題のように $|B_i|^2$ の値は小さくなるが、i が多くなると i が増加しても不能問題のようになつて $|B_i|^2$ の値は小さくならない。

注***　$V_i \cdot V_j = \delta_{ij}$　　$\begin{cases} i = j & \delta_{ij} = 1.0 \\ i \neq j & \delta_{ij} = 0.0 \end{cases}$

$W_i \cdot W_j = \delta_{ij}$